心理科学进展 2020, Vol. 28, No. 10, 1662–1677
Advances in Psychological Science

https://dx.doi.org/10.3724/SP.J.1042.2020.01662

多感觉整合中的声音诱发闪光错觉效应*

王爱君[1] 黄 杰[1] 陆菲菲[1] 何嘉滢[1] 唐晓雨[2] 张 明[1]

([1]苏州大学心理学系，心理与行为科学研究中心，苏州 215123)
([2]辽宁师范大学心理学院，儿童青少年健康人格评定与培养协同创新中心，大连 116029)

摘 要 声音诱发闪光错觉效应是典型的视听整合错觉现象，是指当视觉闪光刺激与间隔100 ms内的听觉声音刺激不等数量呈现时，被试知觉视觉闪光的数量与听觉声音的数量相等。声音诱发闪光错觉的影响因素既包括自下而上和自上而下的被试内差异因素，也包括视听刺激依赖程度、视听整合的发展和视听刺激知觉敏感性等被试间差异因素。该效应的产生在时程上主要体现在早期加工阶段，在脑区上主要涉及多处皮层及皮层下相关脑区。未来研究应进一步考察注意、奖赏和视听整合方式等认知加工对声音诱发闪光错觉的影响，同时也应该关注声音诱发闪光错觉对记忆和学习的影响以及结合计算模型和神经科学的手段进一步探讨其认知神经机制。

关键词 声音诱发闪光错觉；融合错觉；裂变错觉；多感觉整合
分类号 B842

1 引言

我们对外界的感知来自于多个感觉系统，不同感觉通道之间的刺激输入会发生交互作用，多个感觉系统之间的整合与竞争使得我们更有效的对外界做出反应。多感觉整合是(multisensory integration)是个体将不同感觉通道(如视觉、听觉、触觉)的信息整合为统一、连贯、稳定的有意义知觉过程(Stein & Stanford, 2008; 文小辉，李国强，刘强，2011)。多感觉整合的一种表现形式为多感觉错觉效应(multisensory illusion)，典型例子如麦格克视觉主导效应(McGurk effect)和声音诱发闪光错觉听觉主导效应(Sound-induced flash illusion)。麦格克效应和声音诱发闪光错觉均可以被认为是跨感觉通道的双稳态知觉现象(bistable perception)，即当我们面对同样的刺激产生两种不同的、相互竞争的知觉(Sterzer, Kleinschmidt, & Rees, 2009; Wang, Arteaga, & He, 2013)。虽然在多数双通道交互作用的情况下，视觉通道占主导。但在时间维度上，听觉通道通常占主导地位，比如声音诱发闪光错觉效应(Shams, Kamitani, & Shimojo, 2000, 2002)和时间口技效应(temporal ventriloquism) (Bertelson & Aschersleben, 2003)。

声音诱发闪光错觉是Ladan Shams提出的一种典型的以听觉信息为主导的视听整合现象，是指当视觉闪光伴随不相等数量的听觉声音在100 ms内相继或同时呈现时，个体错觉性的将视觉闪光的数量与听觉声音的数量知觉为相等(Shams et al., 2000, 2002)。该错觉效应被认为是多感觉整合中的经典错觉现象，同时也被认为是跨感觉通道的双稳态知觉现象。Shams等人(2000, 2002)首次采用声音诱发闪光错觉范式进行研究，提出了在视听信息整合的过程中听觉信息可以主导视觉信息的加工，即产生多感觉整合的错觉效应。研究采用两个间隔为57 ms的听觉声音并伴随一个与第一个听觉声音间隔23 ms的视觉闪光快速呈现，然后让被试判断知觉到的视觉闪光个数。结果发现，大部分被试会知觉到两个视觉闪光，这种现象通常也被称为裂变错觉(Fission illusion) (Shams et al., 2000, 2002)。随后，Anderson, Tiippana和Sams (2004)研究采用Shams的研究范式，设置多个视

收稿日期：2019-11-20
* 国家自然科学基金(31700939, 31871092, 31600882)、江苏省基础研究计划(BK20170333)资助。
通信作者：张明, E-mail: psyzm@suda.edu.cn;
唐晓雨, E-mail: tangyu-2006@163.com

觉闪光刺激伴随一个听觉刺激，让被试报告视觉闪光个数。结果发现，当两个视觉闪光伴随一个听觉声音时，两个视觉闪光会被错误知觉为一个视觉闪光，这种现象通常也被称为融合错觉(Fusion illusion) (Abadi & Murphy, 2014; Andersen, Tiippana, & Sams, 2004)。

关于声音诱发闪光错觉效应产生的原因主要归为两类，其一是由于个体对外界刺激输入的主观判断导致了其主观知觉闪光刺激数量的改变(Andersen et al., 2004; Shams et al., 2002)；其二为个体视知觉敏感性和判断标准的降低(McCormick, & Mamassian, 2008)。造成个体主观判断和知觉敏感性变化的一个最为重要的原因是听觉声音刺激的出现对视觉闪光刺激判断产生的因果关系知觉，即多感觉整合中的因果关系(Shams & Beierholm, 2010)。多感觉整合中因果关系知觉的研究发现，同步呈现的无任何提示意义但与视觉运动刺激方向一致的听觉刺激能够促进视知觉的加工(Kim, Peters, & Shams, 2011)。对于声音诱发闪光错觉效应，Shams, Ma 和 Beierholm (2005)采用经典的声音诱发闪光错觉范式考察听觉声音刺激的出现对视觉闪光刺激判断产生的因果关系。结果发现，听觉声音刺激的个数能够影响被试对于视觉闪光刺激的知觉，并且这一过程符合贝叶斯因果推论模型(Shams et al., 2005)。

声音诱发闪光错觉效应一直是视听整合研究中的热点问题，自 Shams 等人(2000)提出该效应以来，大量研究开始关注该效应的影响因素和神经机制等问题。但目前缺乏全面且系统的对声音诱发闪光错觉效应进行综述。本文尝试对声音诱发闪光错觉进行全面地综述。首先，从多角度出发阐述影响声音诱发闪光错觉的相关因素；然后，从多种神经科学技术手段阐述声音诱发闪光错觉的认知神经机制；最后，提出未来研究的展望。

2　声音诱发闪光错觉的影响因素

2.1　影响声音诱发闪光错觉被试内差异的因素

声音诱发闪光错觉的错觉量会受到某些因素的影响而发生改变，即在实验设计中，不同实验条件之间的错觉效应不同。综合以往的研究发现，造成声音诱发闪光错觉被试内的因素主要有自下而上的物理刺激因素和自上而下的认知因素两大类。

2.1.1　自下而上的物理刺激因素

刺激的空间特征可以影响个体对视听信息的加工效果，进而造成声音诱发闪光错觉的被试内差异。Lewald 和 Guski (2003)的研究发现，被试判断闪光刺激的个数既取决于时间上的差异，也取决于空间上的差异。研究表明，两种刺激在时间和空间上的接近对于体验因果关系至关重要(Lewald & Guski, 2003)。Innes-Brown 和 Crewther (2009)在 Shams 等人(2002)的经典研究的基础上加入了视听刺激同时呈现在同侧位置以及呈现异侧位置的条件，考察了视听刺激空间不一致性对声音诱发闪光错觉的影响。结果表明，空间不一致性对声音诱发闪光错觉的产生没有影响，但当视听刺激空间一致时，知觉错觉敏感性逐步提高(Innes-Brown & Crewther, 2009)。Máté, Anette, Verena, Masataka 和 Uta (2015)进一步考察了空间一致性的作用。研究将视觉、听觉刺激与动态连续闪烁抑制(dynamic continuous flash suppression, DCFS)相结合，对两只眼睛呈现不同的视觉刺激。其中对遮挡眼睛呈现一组三个动态蒙德里安，对目标眼睛呈现三个灰色的圆盘，位置分别为：左、中和右，当其中一个圆盘呈现闪光时，听觉刺激呈现在目标眼睛对应的屏幕一侧。结果表明，同时发出的声音可以增强没有意识到的视觉信号，并且这一过程依赖于听觉和视觉信号的空间一致性。此外，同时发出的声音会根据闪光的可见度影响闪光的位置。对于那些被判断为不可见的闪光，声音对于报告的闪光位置的空间偏差最强(Máté et al., 2015)。Abadi 和 Murphy (2014)在 Shams 等人(2002)范式的基础上，增添了空间位置的变化，即视觉和听觉刺激运动方向相同(闪光从左边运动到右边，声音也是依次从左向右，为一致性刺激)与视觉和听觉刺激运动方向不同(闪光从左边运动到右边，声音从右向左，为不一致刺激)。结果表明，裂变错觉比融合错觉效应量更大，且证明了视听刺激组合(空间一致性)对声音诱发闪光裂变和融合错觉均产生了影响(Abadi & Murphy, 2014)。DeLoss 和 Andersen (2015)在以往研究的基础上，增大了研究中刺激的空间差异，即将以往 20°空间差异增加为 50°。结果发现，空间差异对声音诱发闪光错觉无影响(DeLoss & Andersen, 2015)。

刺激强度的差异会影响被试的反应，那么感

觉通道间信号强度的差异可能影响声音诱发闪光错觉效应的错觉量大小。Kawabe (2009)的研究改变了视觉闪光的亮度(增加视觉闪光亮度), 考察是否能够通过变化视听刺激的时间间隔产生融合错觉。结果发现, 当视听刺激时间间隔接近时, 声音更容易捕获闪光, 导致融合错觉, 而且被试知觉到的闪光亮度是增强的。此外, 当部分声音给予高频率时, 被试知觉到的闪光亮度也是增强的。但当视听刺激时间超过被试捕获的范围, 刺激属性匹配消失。结果说明, 当感觉通道刺激的强度发生变化时可以影响声音诱发融合错觉效应的强度(Kawabe, 2009)。Andersen 等人(2004)在 Shams (2002)研究基础上考察声音属性是否会影响声音诱发闪光错觉。实验加入了两组声音强度刺激, 一组正常声音强度(80dB, 与 Shams 实验参数相同)与一组低水平声音强度(10dB, 接近被试听觉阈限)。结果表明, 裂变错觉在两组不同的声音强度水平没有显著性差异, 即裂变错觉不受声音强度的影响, 但是融合错觉持续出现直到当听觉接近阈限水平, 且融合错觉在低水平声音强度下消失。研究发现, 被试知觉到的融合错觉因声音强度变化而发生变化, 也表现出当外界刺激输入发生改变时, 错觉现象会随着被试对刺激的主观判断而发生改变(Andersen et al., 2004)。研究表明, 裂变错觉较稳定, 不随声音强度变化, 而融合错觉现象稳定性较弱, 容易受到刺激条件变化的影响。此外, Setti 和 Chan (2011)探究了复杂的视觉刺激是否影响声音诱发闪光错觉现象的敏感性。结果表明, 被试在看到熟悉的视觉刺激时, 显著降低了声音诱发闪光敏感性, 即错觉数量减少。研究也从侧面也说明了视觉刺激的强度可以影响声音诱发闪光错觉(Setti & Chan, 2011)。

刺激的时间间隔被认为是多感觉整合中的重要影响因素(Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010)。因此, 视听刺激的时间间隔可能影响个体对于视听整合信息的加工, 进而造成错觉量的个体内变异。Shams 等人(2002)呈现一个视觉闪光刺激伴随多个听觉声音刺激, 要求被试对视觉闪光刺激数量进行判断。研究发现, 声音诱发闪光错觉是被试基于自己的主观视知觉, 并非由于任务难度或知觉偏向所造成。并且一个强裂变错觉的产生需要两个声音刺激的时间间隔低于 70 ms。而裂变错觉减弱, 则需要两个声音刺激的时间间隔大于 150 ms。研究说明了两个声音刺激的时间窗可以影响裂变错觉知觉(Shams et al., 2002, 2005)。Zhang 和 Chen (2006)使用 fMRI 技术考察裂变错觉引起的视觉和听觉之间的跨通道交互作用。结果发现, 当听觉和视觉刺激之间的刺激间隔延迟为 25 ms 时, 视觉皮层的功能磁共振成像信号对裂变错觉的反应明显高于真实的单闪光。当延迟时间延长到 300 ms 时, 这种信号增强消失了。研究表明, 裂变错觉效应对视觉皮层的脑活动具有显著的影响, 这种影响程度对刺激间延迟具有动态敏感性(Zhang & Chen, 2006)。近期, Kostaki 和 Vatakis (2016)进一步采用不同时间间隔的声音诱发闪光错觉范式考察不均等数量的感觉通道刺激输入之间的竞争。结果发现, 裂变错觉在不同 SOA 下会产生不同的错觉强度(Kostaki & Vatakis, 2016)。

总体而言, 声音诱发闪光错觉效应一方面容易受到物理刺激因素影响而产生个体内变异, 另一方面又具有较强的稳定性, 即不容易消失, 特别是裂变错觉, 其不会因为视听刺激属性的变化减弱或消失, 也说明了裂变错觉是一种稳定的错觉现象, 且当声音的时间间隔较近时或当视觉刺激呈现在外周视野时, 裂变错觉效应最大。但是融合错觉在部分研究中没有发现, 可能由于研究中实验条件、实验设备等影响, 并且融合错觉会受到刺激物理属性(如声音强度)的影响使之减弱或消失; 当视听刺激的时间间隔较短时, 更容易发生融合错觉现象。因此, 相对裂变错觉而言, 融合错觉效应较弱且不稳定。

2.1.2 自上而下的认知因素

刺激物理属性的改变对声音诱发闪光错觉存在影响。但即使面对同样的物理刺激, 个体的认知状态不同, 也可能造成声音诱发闪光错觉的错觉量不同, 也就是双稳态知觉过程中, 尽管自下而上的刺激保持不变, 但产生与不产生错觉在感知过程中交替出现。此外, 相对于物理刺激这种自下而上的影响因素, 自上而下的认知因素在生活中更为常见。目前已有的研究主要围绕注意、反馈方式和认知期望等因素进行了研究。

Mishra, Martínez 和 Hillyard (2010)提出声音诱发视觉错觉不仅是跨通道整合的结果, 应该还有注意分配的作用。研究采用 ERP 技术, 在 Shams 等人(2002)范式基础上增加了刺激的空间位置(高

视野和低视野)与是否注意(注意和不注意)两个变量，要求被试在高视野或低视野判断视觉闪光刺激数量。研究表明，在低视野刺激中，PD120/PD110 成分位于腹侧枕、颞外纹视觉皮层，而随后的 PD180 和 ND250/ND240 成分位于颞上回。而枕颞皮层和颞上回被认为是多感觉通道信息产生交互的脑区。这些 ERP 成分与之前 Mishra, Martinez, Sejnowski 和 Hillyard (2007)研究发现的声音诱发闪光错觉的神经机制的成分有高相似性，并且在此研究中加入注意分配使得这些成分均被增强。这些结果证明了注意分配对跨通道交互与声音诱发闪光错觉的关系存在影响，说明了裂变错觉并不全是自动整合加工的结果，还有注意分配参与了视觉和听觉信号的整合(Mishra et al., 2010)。除了注意分配能够影响声音诱发闪光错觉之外，内源性注意也能够影响声音诱发闪光错觉效应。Georgios 和 Julian (2018)的研究考察了内源性注意对声音诱发闪光错觉的影响。结果发现，在高工作记忆负荷下，被试感知到的闪光错觉量增加，且与反应偏差无关。证明了视听整合可以被认知资源的数量调节，为非言语视听刺激的整合在注意资源减少的情况下得到增强提供了证据，因此支持了自上而下的注意控制在多感觉整合中发挥重要作用(Georgios & Julian, 2018)。同样地，于薇，王爱君和张明 (2017)采用经典的声音诱发闪光错觉的范式，通过操纵注意资源的分配方式考察了主动注意听觉通道的声音刺激对声音诱发闪光错觉产生的影响。结果发现，裂变错觉受到注意资源分配程度的影响，但融合错觉不会受到注意资源分配的影响(于薇等, 2017)。近期，张明等人(2018)将注意以内源性的方式指向视觉通道或者听觉通道，考察了基于通道的内源性注意对声音诱发闪光错觉的影响。结果发现，当注意内源性地指向视觉通道时，裂变错觉的量显著减小；当注意内源性地指向听觉通道时，裂变错觉的量存在增大的趋势。说明了基于通道的内源性注意可以影响声音诱发闪光错觉中的裂变错觉，但不影响融合错觉(张明等, 2018)。

Rosenthal, Shimojo 和 Shams (2009)采用 Shams 等人(2002)的研究范式，在被试完成任务时给予被试任务判断正确与否的反馈，探讨了反馈方式是否可以影响声音诱发闪光错觉现象。结果表明，无论给予被试正确与否的反馈，均不能显著减小或者消除这种错觉现象，即这种声音诱发闪光错觉现象较为稳定。但是，当给予被试货币奖励(即依据被试结果的正确率给予相应的奖励)时，这种错觉性的主导效应却有下降的趋势(Rosenthal et al., 2009)。van Erp, Philippi 和 Werkhoven (2013)进一步认为，既然脑成像上可以发现错觉闪光和真实闪光在视觉皮层的激活有差异，那么在反馈训练中，虽然反馈和没有反馈现象无显著差异，但反馈在一定程度上降低知觉到的错觉。因此，他们认为以往的研究忽略了真实闪光和错觉闪光在对比度、强度和持续时间上差异，并错误地假定真实和错觉闪光有相似知觉而且不能区分。因此，他们提出增加反应类型，即在反应一个闪光和两个闪光的基础上，增加了“某些事物不同于一个或两个闪光”，记为反应类型 3。结果发现，被试在按键反应时，选择反应类型 3 中有 44%的是错觉闪光，有 4%是真实闪光，表明可能真实闪光和错觉闪光在强度、对比度和持续时间等特性上存在差异，而被试可以通过自己来判断出真实闪光。此外，冲突探测(conflict detection)也可能会影响被试的判断反应。因为被试在进行判断时，可能没有区分真实和错觉闪光之间的定性差异，而是可能注意到了视觉脉冲和听觉脉冲之间的“冲突”。在匹配条件中，闪光和声音脉冲的时间和数量是相同的。而在错觉闪光试次中，闪光和声音脉冲的时间(视听刺激时间间隔 23 ms)和数量(声音数量多于或少于闪光数量)是不同的。因此即使在指导语中说明被试反应只依据被试视知觉，但被试对这种冲突探测却体现在被试反应上。这意味着被试并非依据错觉和真实闪光的定性特征，而是依据视觉和听觉信息之间的冲突进行结果判断(van Erp et al., 2013)。以上研究表明，正确的结果反馈并没有显著影响声音诱发闪光错觉现象。因此，可以从实验结果推断，错误的结果反馈也不会显著影响声音诱发闪光错觉，并且虚假反馈可能会增加被试报告裂变错觉的比例。通过 van Erp 等人(2013)的实验可以发现，被试自身可能可以区分真实闪光和错觉闪光，但具体被试是依据什么判断以及是否存在冲突探测的影响都有不确定性，还需要进一步的研究。综上这些研究表明，直接结果反馈(即说明被试判断正确与否)对声音诱发闪光错觉没有显著影响，但当增加结果反馈类型时(即反应类型增加)，可以改变被试知觉到

的闪光错觉。

除了注意和反馈方式，还有研究探讨了自上向下的认知期望对声音诱发闪光错觉的影响。研究利用改变指导语与实验试次比例来操控认知期望，实验 1 和实验 2 采用的三个指导语条件一致(“请判断闪光个数”、“含有两个闪光的条件为 80%, 判断闪光个数”和“含有一个闪光的条件为 80%, 判断闪光个数”)，实验试次比例不同；实验 3 中采用三个相同指导语(“请判断闪光个数”、“所有实验条件出现的概率相等，判断闪光个数”和“所有实验条件出现的概率相等，判断闪光个数)，实验试次比例与实验 1 相一致。结果表明，当指导语的提示比例与试次的实际比例相一致时，可以显著提高观察者的判断准确性，反应时变快；当指导语的提示比例与试次的实际比例不一致时，指导语依然可以影响观察者在判断裂变错觉时的准确性和反应时。研究表明，自上而下的认知预期可以显著减少裂变错觉，加速被试的判断。而在融合错觉中这种影响并不明显，可能是由于融合错觉相对不稳定性所导致的(Wang et al., 2019)。

综上所述，在声音诱发闪光错觉的被试内变异研究中，研究者既关注了自下而上的物理刺激因素对声音诱发闪光错觉的影响，同时也关注了一些认知因素，特别是自上而下的认知因素对声音诱发闪光错觉的影响。物理刺激因素更多地体现了声音诱发闪光错觉现象的产生是由于在不同的物理刺激条件下被试视知觉敏感性降低(McCormick & Mamassian, 2008)。例如，Kumpik, Roberts, King 和 Bizley (2014)的研究通过降低视觉刺激的空间可靠性探讨视觉知觉敏感性对声音诱发闪光错觉的影响。研究发现，声音诱发闪光错觉在外周视野更常见，且被试知觉到的裂变错觉数量大于融合错觉数量，说明声音诱发闪光错觉是被试对视野范围呈现闪光的知觉敏感性决定。自上而下的认知因素则更多地体现了声音诱发闪光错觉现象的产生是由于被试对外界刺激输入的主观判断导致了被试主观知觉闪光刺激数量的改变。例如，Shams 等人(2002)研究认为，声音诱发闪光错觉是被试基于的主观视知觉，并非由于任务难度或物理刺激不同造成的知觉偏向所造成(Kumpik et al., 2014)。

2.2 影响声音诱发闪光错觉被试间差异的因素

声音诱发闪光错觉效应中，不同被试的之间存在错觉量的差异，不同组别的被试之间也存在声音诱发闪光错觉的差异。Mishra 等人(2007)的研究发现，这种错觉现象在个体间的差异是较大的，闪光错觉发生的比例范围在个体间是 3%~86%。下面将分成三个相关因素进行阐述，主要包括对视听信息依赖程度、视听整合发展以及视听刺激知觉敏感性。

2.2.1 视听刺激依赖程度

声音诱发闪光错觉的个体差异可能来自个体对视觉或听觉信息依赖程度的差异，即对视觉信息依赖程度高的个体不易受到听觉信息的影响，进而错觉量较小；对听觉信息依赖程度高的个体容易受到听觉信息影响，进而错觉量也较大。一项研究通过对比了音乐家和非音乐家对于声音诱发闪光错觉这种视听信息整合之间的差异。研究发现，受过音乐训练的人能够更快更准确地加工视听线索，并且他们声音诱发闪光错觉的能力增强以及视听信息绑定的时间窗口更短(Bidelman, 2016)。这可能是因为音乐家通过长期训练使得听觉信息的加工能力增强。另一项研究表明，相比于正常双眼被试，早年失去一只眼睛功能的单眼被试声音诱发闪光错觉的错觉量减小(Moro & Steeves, 2018)，说明了视听整合中出现的适应补偿现象是由于单眼被试的视觉输入的降低所引起的。以往的研究表明，当单一感觉信息不可靠时，人脑可以通过整合多感觉信息来增强对外界的感知，即当知觉单一感觉信息的能力下降时，多感觉信息对知觉的影响更大。对于单眼被试来说，当视觉减弱时，听觉可以提供关于环境特性的信息(Klatzky, Marston, Giudice, Golledge, & Loomis, 2006)。

对特殊人群(联觉者、偏头痛患者、多发性硬化患者、弱视者、自闭症谱系障碍患者、轻度认知障碍被试和精神分裂症患者)的声音诱发闪光错觉研究也支持上述观点，即对视听信息依赖程度上的差异可能影响声音诱发闪光错觉的个体间差异。Whittingham, Mcdonald, & Clifford (2014)的研究发现，联觉者随着年龄的增长，声音诱发闪光错觉的错觉量减小(Whittingham et al., 2014)。Brighina 等人(2015)的研究发现，对于偏头痛患者，特别是在发作期间，当呈现一个视觉闪光与两个听觉声音结合时，裂变错觉减小甚至消失(Brighina et al., 2015)。Yalachkov 等人(2019)的研究发现，多

发性硬化患者感知到的闪光错觉量多于正常健康被试(Yalachkov et al., 2019)。Narinesingh, Goltz 和 Wong (2017)的研究发现, 对于双眼观察且听觉声音刺激呈现在前的裂变错觉, 随着刺激呈现的不同步性增加, 正常被试表现出错觉强度降低, 而弱视患者的错觉强度则保持不变, 即弱视组的时间绑定窗口更宽。对于视觉闪光刺激呈现在前的融合错觉, 弱视个体错觉强度降低(Narinesingh et al., 2017)。Bao, Doobay, Mottron, Collignon 和 Bertone (2017)的研究发现, 自闭症谱系障碍患者和正常发育个体都对裂变错觉具有敏感性, 但是自闭症谱系障碍患者更容易出现融合错觉(Bao et al., 2017)。Chan 等人(2015)的研究发现, 在较长的听觉时间间隔内, 与健康被试相比, 轻度认知障碍被试感知到更多的错觉(Chan et al., 2015)。Vanes 等人(2016)的研究发现, 精神分裂症患者的融合错觉量显著降低(Vanes et al., 2016)。同样地, Haß 等(2017)通过使用不同时间间隔的刺激诱发裂变错觉, 发现了精神分裂症患者与健康被试在视听整合方面存在差异(Haß et al., 2017)。近期, Gieseler, Tahden, Thiel 和 Colonius (2018)考察了使用助听器的老年人与未使用助听器的轻度听力损失者(非使用助听器者)在声音诱发闪光错觉量上的差异, 并通过改变刺激开始的不同步性来考察相应的整合窗口。结果发现, 轻度听力损伤和助听器的使用都会影响错觉量的大小, 原因可能是因为助听器的使用会逆转听力损失对视听整合的影响(Gieseler et al., 2018)。

总体而言, 不同群体之间的比较体现出了认知加工过程中对视觉和听觉信息的依赖程度能够影响声音诱发闪光错觉效应出现的个体差异。具体表现为, 对视觉信息依赖程度高的个体不易受到听觉信息的影响, 进而导致声音诱发闪光错觉这种听觉刺激驱动的错觉量减小; 对听觉信息依赖程度高的个体容易受到听觉信息影响, 进而导致这种声音诱发闪光错觉这种听觉主导效应增大。然而, 目前不同群体的研究证据除了被试在视听信息依赖程度的差异之外, 还可能会存在其他的影响因素影响错觉量的大小。未来的研究一方面可以考察其他被试群体声音诱发闪光错觉的错觉量大小为视听信息依赖程度这一影响因素提供证据, 更为重要的是, 可以在现有研究的基础上进一步量化和深入考察视听信息依赖程度如何影响这些被试群体的声音诱发闪光错觉。

2.2.2　视听整合的发展

大脑为了有效整合来自不同感觉通道的信息使用了时间绑定窗口。以往研究发现时间绑定窗口的宽度随着年龄的增长而增加, 以便在感觉转换下降时最大限度利用多感觉通道输入的信息。DeLoss, Pierce 和 Andersen (2013)使用声音诱发闪光错觉范式考察声音诱发闪光错觉是否存在年龄差异。结果发现, 与年轻人相比, 老年人抑制无关跨通道信息能力降低, 整合信息能力更强, 即表现出更多的错觉现象, 这一研究表明, 多感觉整合与年龄存在分离效应, 老年人更容易受到多感觉整合的影响(DeLoss et al., 2013)。McGovern, Roudaia, Stapleton, McGinnity 和 Newell (2014)的研究发现, 老年人即使在刺激的时间间隔(SOA)较大的情况下也会产生裂变错觉。但是对于融合错觉, 老年人和年轻人之间的错觉量并不存在显著差异(McGovern et al., 2014)。同样地, Chan 等人(2017)比较了年轻人和老年人声音诱发闪光错觉任务的错觉量。结果发现, 在较长的 SOA 中老年人比年轻人更容易产生声音诱发闪光错觉(Chan et al., 2017)。随后, Chan, Connolly 和 Setti (2018)考察了年轻人与老年人在面对长(±70、±110、±150、±190 和±230)和短(±70、±150 和±230)两种不同 SOA 时, 是否对声音诱发的闪光错觉表现出不同的易感性。结果发现, 这两组被试在小范围的不同步性时, 对错觉表现出更高的易感性, 但是仅限于较长的时间窗。另外, 研究也发现改变 SOA 的量确实会影响对感知到错觉量的敏感性(Chan et al., 2018)。

近期, Hernández, Setti, Kenny 和 Newell (2019)的研究提供了关于声音诱发闪光错觉的第一个横向大数据研究的证据。研究以 3955 名 50 岁以上的成年人为样本, 考察了年龄、认知状况和性别对声音诱发闪光错觉敏感性的影响。结果发现, 声音诱发闪光错觉敏感性的研究结果与先前的研究一致, 认为老化在多感觉加工和增大的时间绑定窗口中起作用, 即年龄越大、蒙特利尔认知评估(MoCA)分数越低, 声音诱发闪光错觉的敏感性越高(Hernández et al., 2019)。O’Brien, Ottoboni, Tessari 和 Setti (2017)采用声音诱发闪光错觉任务考察老年人的视听整合能力, 老年人在运动前后(持续时间在 60 到 80 分钟之间)接受测试。结果发

现，健康并经常锻炼的老年人显示出更有效的多感觉加工能力(O'Brien et al., 2017)。Merriman, Whyatt, Setti, Craig 和 Newell (2015)考察了平衡功能的改善是否与老年人多感觉加工能力的改善有关，尤其是那些过去有跌倒经历的老年人。研究对 76 名健康且有跌倒倾向的老年人进行平衡训练，并使用声音诱发闪光融合错觉任务考察多感觉加工能力是否改善。结果发现，易跌倒老年人的平衡功能的改善和多感觉整合的改善之间存在相关。表明了平衡训练干预可以增强老年人的身体姿势控制能力，对易跌倒的老年人产生更有效的多感觉加工(Merriman et al., 2015)。

上述研究表明老年人和年轻人视听整合的时间窗大小影响了声音诱发闪光错觉的效应量，那么什么原因造成了两组被试群体视听整合的时间窗的差异？Cecere, Rees 和 Romei (2015)发现声音诱发闪光错觉现象的视听刺激整合时间窗是 100 ms 左右，而 α 频段的振荡活动周期也是 100 ms 左右，因此他们认为，α 频段的振荡周期可以调节视听事件整合的时间窗。研究也发现，个体 α 频率(individual alpha frequency, IAF)峰值和错觉时间窗的大小之间呈正相关关系。此外研究也发现，错觉的时间窗随 α 频率的振荡活动而变化。说明了 α 振荡可以调节闪光错觉的整合时间窗。因此，Cecere 等人认为，闪光错觉是在快速视觉刺激加工的关键时间窗内由连续的声音诱发视觉皮层兴奋性导致的突然变化所引起的现象(Cecere et al., 2015)。Keil 和 Senkowski (2017)同样通过脑电研究考察了个体 α 频率与个体知觉闪光错觉的关系。结果在验证 Cecere 等人(2015)的基础上，扩展了原有结果。表明了个体 α 波频率与闪光错觉的错觉率呈负相关，即低 IAF 个体视听觉错觉更强。并且源分析表明视觉皮层，特别是距状沟参与表征(Keil & Senkowski, 2017)。

综上所述，个体多感觉整合能力越强，声音诱发闪光错觉量越大。相比于年轻人，老年人声音诱发闪光错觉中的裂变错觉和融合错觉效应均更大(Sun et al., 2020)。之所以表现多感觉错觉效应增大的原因在于老年人多感觉整合能力可能比年轻人更强。例如，以往关于多感觉整合年龄效应的研究表明老年人的多感觉整合能力强于年轻人(Mahoney, Li, Oh-park, Verghese, & Holtzer, 2011; DeLoss et al., 2013)。关于老年人多感觉整合效应增强的原因 Laurienti, Burdette, Maldjian 和 Wallace (2006)认为可能是由于补偿机制，即多感觉通道信息的加工能力较强能够补偿单一感觉通道信息加工能力的不足(Laurienti et al., 2006)。以往研究表明，当单一感觉通道信息不可靠时，人脑可以通过整合多感觉通道的信息来增强对外界的感知，即当知觉单一感觉通道信息的加工能力下降时，多感觉通道信息的加工对个体的知觉过程影响更大。对于老年人来说，当视觉通道信息减弱时，听觉通道信息可以作为补偿提供更多关于环境特性的信息(Klatzky et al., 2006)。

2.2.3 视听刺激知觉敏感性

声音诱发闪光错觉是视听整合的一种现象，被试对视觉刺激和听觉刺激加工过程中早期感知觉阶段的差异也可能影响错觉效应的大小。有研究在 Shams 等人(2002)研究的基础上，考察感知觉敏感性对声音诱发闪光错觉的影响。结果发现，声音诱发闪光错觉的错觉量受到知觉敏感性的影响，即知觉敏感性越低，闪光错觉量越大(McCormick & Mamassian, 2008)。Kumpik 等人(2014)考察了听觉刺激影响视觉感知可能性的空间因素。研究者通过操纵视觉与听觉刺激的空间一致性、声音定位信号的类型、视觉刺激的离心度和大小，考察视觉知觉敏感度在多大程度上决定了闪光错觉的出现。信号检测论结果发现，声音诱发闪光错觉在外周视野更常见，且被试知觉到的裂变错觉量大于融合错觉量(Kumpik et al., 2014)。

鉴于知觉敏感性能够影响声音诱发闪光错觉的错觉量，de Haas, Kanai, Jalkanen 和 Rees (2012)的研究进一步提出了人类大脑结构的差异造成了个体对声音诱发闪光错觉敏感性的差异。研究结果发现，当初级视觉皮层局部(BA17&18)的灰质体积较小时，被试更容易产生错觉，并且无论视觉刺激呈现在上视野还是下视野，灰质体积与被试的闪光错觉量均呈负相关。研究表明，个体大脑初级视觉皮层的灰质体积与声音诱发闪光错觉有关(de Haas et al., 2012)。此外，已有研究证明伽马波段(N30 Hz)在多感觉信息加工中起着重要作用，颞上沟的 γ 氨基丁酸(GABA)神经传递有助于产生 γ 带振荡(gamma band oscillation, GBO)，其可通过激活代谢型谷氨酸受体来维持(Balz, Roa Romero et al., 2016)。因此，Balz 等人(2016)提出 GABA 和谷氨酸系统的差异可能会导致多感觉信

息加工的个体差异。通过 EEG 和 MRI 两种研究方法, 考察了 GABA 和颞上沟(STS)中的谷氨酸浓度、局部 GBO 和声诱发闪光错觉的错觉量之间的关系。研究发现, GABA 水平通过其对 GBO 的调节塑造视听感知的个体差异, STS 中 GABA 浓度与 GBO 功率之间存在很强的正相关关系, GABA 浓度与个体感知闪光错觉的可能性之间存在正相关关系(Balz, Keil et al., 2016)。

综上所述, 视觉和听觉通道中刺激知觉敏感性能够影响声音诱发闪光错觉, 表现为视觉敏感性降低, 闪光错觉量越大。并且研究发现这种敏感性与大脑结构的差异存在关系。鉴于刺激的知觉敏感性能够影响声音诱发闪光错觉, 近期 Sun 等人(2020)通过在声音诱发闪光错觉经典范式之前重复呈现听觉通道的刺激, 考察听觉通道刺激的知觉敏感性变化如何影响声音诱发闪光错觉。研究结果发现, 听觉通道刺激的重复呈现可以使得听觉声音刺激的敏感性下降, 进而使得错觉量降低(Sun et al., 2020)。因此, 被试对于视觉和听觉知觉敏感性性的改变也被认为是声音诱发闪光错觉现象产生的主要原因(Kumpik et al., 2014; McCormick & Mamassian, 2008)。目前现有的研究均从单一的视觉通道刺激和听觉通道刺激的知觉敏感性变化考察其如何影响声音诱发闪光错觉, 未来的研究可以从视觉刺激和听觉刺激知觉敏感性相对变化的角度考察声音诱发闪光错觉的错觉量大小如何受到刺激知觉敏感性的影响。

3 声音诱发闪光错觉的认知神经机制

3.1 声音诱发闪光错觉的时程机制

大脑在接受视听信息的输入后, 开始对其进行整合加工。那么大脑在接受声音诱发闪光错觉任务中的视听刺激后的不同阶段是如何加工这些视听刺激而产生不同的错觉? 研究者通过具有高时间分辨率的脑电技术(EEG 和 ERP)和脑磁图技术(MEG)针对这一问题进行了探讨。研究结果表明, 声音诱发闪光错觉中的视听觉信息整合发生在认知加工的早期阶段。

Shams, Kamitani, Thompson 和 Shimojo (2001)使用 ERPs 通过视觉诱发电位(VEPs)探究了裂变错觉这种跨通道整合现象对视知觉产生的影响是否在通道特定的(modality-specific)视觉通路。研究对比了有声音出现的 VEPs 和没有声音呈现的 VEPs。结果发现, 这些跨通道效应发生在枕叶皮层。更重要的是, 裂变错觉闪光的 VEPs 与真实闪光刺激的 VEPs 相似。结果表明, 在视觉皮层中产生的这两种知觉现象有着类似的大脑机制(Shams et al., 2001)。此外, Shams 等人(2002)采用 γ 频段(gamma band) (> 30 Hz)振荡反应来进一步探究声音诱发闪光错觉现象与无错觉现象的神经基础。ERPs 结果发现, 有错觉闪光现象出现时 γ 频段振荡显著高于没有错觉现象发生的试次, 并且视听交互作用只发生在错觉试次中。因此, 这些结果表明了声音可以调节视觉皮层区域的加工, 而调节的强度可以影响最终知觉输出(Shams et al., 2002)。同样地, Shams 等(2005)使用 MEG 考察了声音对视觉的调节及机制, 结果发现, 枕顶叶皮层位置在视觉刺激开始的 35~65 ms 就已经受到了调节, 在刺激后不久(~150 ms), 枕叶、顶叶以及前部区域的活动也被调节(Shams et al., 2005)。Mishra 等人(2007)的研究发现, 裂变错觉在第二个听觉声音呈现 30~60 ms 后产生了对视觉皮层活动幅度的早期调节(Mishra et al., 2007)。随后 Mishra, Martinez 和 Hillyard (2008)的研究发现, 融合错觉在第二个闪光呈现 80~112 ms 后产生了对视觉皮层活动幅度的早期调节(Mishra et al., 2008)。此外, Balz 等(2016)对精神分裂(SCZ)组和健康(HC)组进行了脑电图记录, 考察了声音诱发闪光错觉中的神经振荡和 ERPs。结果发现, 与单一视觉刺激同时出现的多个听觉刺激可以诱发多个视觉刺激的错觉。ERPs 结果显示, 与 HC 相比, 刺激开始后约 135 ms SCZ 的振幅减小, 多感觉信息加工产生改变。此外, 与 HC 相比, 神经振荡的分析表明 SCZ 的枕叶区域在 100~150 ms 后变化为 25~35 Hz (Balz, Roa Romero et al., 2016)。近期, Kaiser, Senkowski, Busch, Balz 和 Keil (2019)的研究发现, 在刺激开始前 0.17~0.05 s, 枕叶电极在 25~41 Hz 之间功率的增加可以预测随后出现的声音诱发闪光错觉。在错觉试次和非错觉试次之间, 对显著性聚类的功率值进行 t 检验, 结果显示在错觉试次中功率更高。研究表明, 单次试验振荡活动可以预测多感觉信号的整合。表明了所观察到的功率调节预先处理了视听信号的整合, 为刺激前振荡在多感觉加工中的支配作用提供了进一步证据(Kaiser et al., 2019)。

3.2 声音诱发闪光错觉的脑区

大脑在加工声音诱发闪光错觉范式中的视听

觉信息时，一个重要的问题是哪些脑区参与了加工？研究者采用高空间分辨率的功能性磁共振成像(fMRI)、经颅直流电刺激(tDCS)、经颅磁刺激技术(TMS)和脑磁图(MEG)等技术针对这一问题进行研究。研究结果表明，声音诱发闪光错觉中的视听觉信息整合可以发生在枕叶视觉皮层、颞上沟、上丘、前额叶和小脑等广泛区域均参与加工了声音诱发闪光错觉。

Watkins, Shams, Tanaka, Haynes 和 Rees (2006)通过视网膜皮层映射图(retinotopic mapping)的方法考察了声音诱发裂变错觉是否可以发生在早期的视觉皮层区域。结果表明，声音改变了被试知觉，反映的是被试的主观知觉而不是物理刺激，并发现裂变错觉在初级视觉皮层(V1)有着更高的激活水平(Watkins et al., 2006)。相反，他们关于融合错觉的研究发现，错觉性视觉闪光在 V1 皮层有着更低的激活水平(Watkins, Shams, Josephs, & Rees, 2007)。此外，经过全脑分析后发现，被激活的大脑区域还包括颞上沟(superior temporal sulcus)、上丘(superior colliculus)。同样地，Zhang 和 Chen (2006)采用动态性的 fMRI 也发现了声音诱发裂变错觉发生在早期的视觉皮层区域，并且效应量的大小与视听刺激间的间隔存在动态性的关联，除了视黄质视觉感知外，大脑外围还可能参与高级别的大脑处理过程(Zhang & Chen, 2006)。随后，Jiang 和 Han (2007)采用 fMRI 技术，基于视知觉的收益与损失来探究声音诱发闪光错觉现象的神经机制。研究发现，视觉闪光伴随听觉声音刺激同时呈现时听觉声音刺激会影响视觉闪光的知觉，当听觉刺激数量大于闪光数量时，被试知觉到了错觉闪光，即知觉收益(perceptual gain)；当听觉刺激数量小于闪光数量时，被试知觉到的闪光数量减少，即知觉损失(perceptual loss)。结果表明，声音诱发闪光错觉现象的变化是双向作用的，既可以产生知觉收益，又可以产生知觉损失。在知觉闪光中错觉知觉的收益激活了缘上回(supramarginal gyrus)、前额叶皮层(prefrontal cortex)和小脑(cerebellum)，这反映出了与工作记忆整合加工相关联的神经网络。而知觉损失发现相关的激活在内侧枕叶皮层(medial occipital cortex)和丘脑(thalamus)，是与早期视觉加工相关的大脑区域(Jiang & Han, 2007)。在磁共振成像中的结构成像中，de Haas 等(2012)考察了声音诱发闪光错觉的个体差异是否与大脑结构中的灰质体积差异有关。结果发现，当初级视觉皮层局部(BA17&18)的灰质体积较小时，被试更容易知觉错觉(de Haas et al., 2012)。

此外，Bolognini, Rossetti, Casati, Mancini 和 Vallar (2011)通过 tDCS 改变(增加或减少)视听刺激交互的可能性。研究发现，通过 tDCS 上调或下调皮层兴奋性可以促进或减少错觉量，并且发现颞、枕皮层区域与声音诱发闪光错觉存在相关(Bolognini et al., 2011)。近期，Maccora 等人(2019)在偏头痛患者的视觉皮层上应用阴极 tDCS，以恢复他们对裂变错觉的生理敏感性。结果发现，tDCS 无法可靠地调节偏头痛患者的声音诱发闪光错觉。原因可能是因为阴极 tDCS 对偏头痛患者皮层兴奋性缺乏作用(Maccora et al., 2019)。Kamke, Vieth, Cottrell 和 Mattingley (2012)使用 TMS 研究发现，角回等顶叶网络在裂变错觉中起到了整合听觉和视觉刺激的作用(Kamke et al., 2012)。Shams 等(2005)使用 MEG 来研究声音对视觉的调节及背后的机制发现，枕顶叶皮层位置在视觉刺激开始的 35~65 ms 就已经受到了调节，在刺激后不久(~150 毫秒)，枕叶、顶叶以及前部区域的活动也被调节(Shams et al., 2005)。Keil, Müller, Hartmann 和 Weisz (2013)使用 MEG 研究发现，在产生错觉之前，左侧颞中回和听觉区域之间的 β 波段相位同步性增强，而视觉区域的相位同步性减弱(Keil et al., 2013)。Chan 等人(2017)在记录 MEG 的同时，要求被试完成声诱发的闪光错觉任务。结果发现，与年轻人相比，老年人在呈现裂变错觉中显示出增强的 β 波带状前刺激，这表明对将要呈现的刺激产生了预测。传递熵分析和前刺激脑磁图数据的动态因果模型显示，与年轻人相比，老年人的听觉和视觉皮层之间的跨通道连接具有更强的与错觉相关的调节(Chan et al., 2017)。

目前对声音诱发闪光错觉的神经机制研究主要基于单一感觉皮层，如初级视觉皮层和听觉皮层。随着研究的不断深入，越来越多的研究发现声音诱发闪光错觉不仅发生在单一感觉皮层，也有多种脑区的交互作用，进一步明确以及发展了声音诱发闪光错觉的神经机制(表 1)。但从表 1 中可以看出，目前关于声音诱发闪光错觉的神经机制研究在实验设计、实验参数等问题上还存在着

表 1　声音诱发闪光错觉效应的认知神经机制研究

研究	错觉类型	研究技术	实验设计 (视觉: A、听觉: V)	重要参数	核心发现
Shams et al., (2001)	裂变错觉	ERPs	V_p、V_f、A_2、V_{p2}、A_2V_{p1}、A_2V_{f1} 外围: P、中央 f	声音刺激: 3.5 kHz 频率, 77 dB 声压级, 呈现 8 ms; 闪光刺激: 呈现在中央或外围 8°, 直径为 2°的白色闪光, 呈现 14 ms; 时间间隔: SOA $_{(A1A2)}$ = 65 ms; SOA $_{(Vp1Vp2)}$ = 67 ms; SOA $_{(A1V1)}$ = 14 ms。	裂变错觉闪光与真实闪光的 VEPs 大脑机制相似
Shams et al., (2002)	裂变错觉	ERPs	实验 1 V_1、V_2、V_3、V_4、A_1V_1、A_2V_1、A_3V_1、A_4V_1、A_1V_1、A_1V_2、A_1V_3、A_1V_4 实验 2 A_2V_1	实验 1 声音刺激: 3.5 kHz 频率, 95 dB 声压级, 呈现 7 ms; 闪光刺激: 呈现在背景为黑色, 中央注视点下方 5°, 直径 2°, 亮度 108 cd/m^2 的白色闪光, 呈现 17 ms; 时间间隔: SOA $_{(A1A2)}$ = 64 ms; SOA $_{(V1V2)}$ = 67 ms; SOA $_{(A1V1)}$ = 23 ms 实验 2 时间间隔: SOA $_{(A1V1)}$ = 0 ms; SOA $_{(A2V1/V1A2)}$ = 25/70/115/160/205/250 ms。	声音可以调节视觉皮层区域的加工, 调节的强度影响最终知觉的输出
Mishra et al., (2007)	裂变错觉	ERPs	A_1、A_2、V_1、V_2、A_1V_1、A_2V_2、A_2V_1、A_1V_2、A_2V_1、null	声音刺激: 76 dB 声压级, 呈现 10 ms 闪光刺激: 亮度 75 cd/m^2 的闪光, 呈现 5 ms; 时间间隔: A1A2、V1V2: SOA = 70 ms; A1V1、A1V1A2V2、A1V1A2、A1V1V2 中 SOA $_{(A1V1)}$ = 10 ms; A1A2V1: SOA $_{(A1V1)}$ = 200 ms。	裂变错觉在第二个声音呈现 30~60 ms 后产生了对视觉皮层活动幅度的早期调节作用
Mishra et al., (2008)	融合错觉				融合错觉在第二个闪光呈现 80~112 ms 后产生了对视觉皮层活动幅度的早期调节作用
Balz, Roa Romero et al. (2016)	裂变错觉	ERPs	A_0V_1、A_0V_2、A_1V_1、A_2V_0、A_2V_2、A_2V_1	声音刺激: 1 kHz 频率, 73 dB 声压级, 呈现 7 ms; 闪光刺激: 呈现在中央注视点下方 4.1°, 直径 1.6°, 亮度 89 cd/m^2 的白色闪光, 呈现 10 ms; 时间间隔: A_2V_1 中 SOA $_{(A1V1)}$ = 0 ms; SOA $_{(V1A2)}$ = 57 ms。	ERPs 结果表明刺激开始约 135 ms 后 SCZ 个体的振幅减小, 神经振荡结果表明 SCZ 被试枕叶区域在刺激开始约 100~150 ms 后变化为 25~35 Hz
Kaiser et al., (2019)	裂变错觉	EEG			刺激开始前 170 秒至 50 秒, 枕叶电极 25 至 41 Hz 之间的功率增加, 可以预测随后出现的错觉感知
Watkins et al., (2006)	裂变错觉	fMRI	V_1、V_2、A_1V_1、A_2V_1、A_1V_2、A_2V_2	声音刺激: 3.5 kHz 频率, 95 dB 声压级, 呈现 10 ms; 闪光刺激: 呈现在背景为灰色(30 cd/m^2), 亮度 420 cd/m^2, 中央注视点下方 8.5~10°的环状带, 呈现 17 ms; 时间间隔: ISI $_{(V1V2)}$ = 56 ms。	声音改变了被试的主观知觉, 裂变错觉可以发生在初级视觉皮层; 同时上丘和右颞上沟区域的活动增强
Watkins et al., (2007)	融合错觉	fMRI		声音刺激: 3.5 kHz 频率, 95 dB 声压级, 呈现 10 ms; 闪光刺激: 呈现在背景为灰色(30 cd/m^2), 亮度 420 cd/m^2, 中央注视点下方 8.5~10°的环状带, 呈现 17 ms; 时间间隔: ISI = 46 ms; SOA $_{(A1A2)}$ = 56 ms; SOA $_{(V1V2)}$ = 63 ms。	融合错觉在视觉皮层为负激活, 在颞上沟、上丘为正激活
Zhang & Chen (2006)	裂变错觉	fMRI	V_1、V_2、A_2V_1	声音刺激: 1.7 kHz 频率, 呈现 7 ms; 闪光刺激: 呈现 17 ms; 时间间隔: ISI $_{(V1V2)}$ = 50 ms; ISI $_{(A1V1)}$ = 25 ms。	当 ISI $_{(A1V1)}$=25 ms 时, 与真实单次闪光相比, 裂变错觉中视觉上皮层中的 fMRI 信号显著增强; 裂变错觉发生在早期视觉皮层, 并且效应量的大小与视听刺激间的间隔存在动态性的关联

续表

研究	错觉类型	研究技术	实验设计(视觉: A、听觉: V)	重要参数	核心发现
Jiang & Han (2007)	闪光错觉	fMRI	V_3、V_5、A_3V_3、A_5V_5、A_5V_3、A_3V_5	声音刺激: 2 kHz 频率, 80 dB 声压级，呈现 7 ms; 闪光刺激：呈现在背景为黑色，中央注视点下方 4.8°，直径 1.6°，亮度 204 cd/m^2 的白色闪光，呈现 17 ms; 时间间隔: A_3V_3、A_5V_5 中 SOA $_{(A1V1)}$ = 23 ms; A_5V_5、A_5V_3 中 SOA $_{(A1A2)}$ = 60 ms; A_3V_3、A_3V_5 中 SOA $_{(A1A2)}$ = 120 ms。	知觉收益时激活左侧缘上回、左侧前额叶皮层和右侧小脑，在知觉损失时激活内侧枕叶皮层和丘脑
de Haas 等 (2012)	裂变错觉	fMRI	V_1、V_2、A_1V_1、A_2V_1、A_1V_2、A_2V_2	声音刺激: 3.5 kHz 频率, 65 dB 声压级，呈现 20 ms; 闪光刺激：呈现在背景为灰色(90 cd/m^2)，中央注视点上方或下方 5°，直径 2°，亮度 140 cd/m^2 的白色闪光，呈现 24 ms; 时间间隔: ISI $_{(V1V2)}$ = 34 ms。	BA17&18 的灰质体积较低时，更容易感知到错觉
Bolognini et al., (2011)	闪光错觉	tDCS	V_1、V_2、V_3、V_4、A_1V_1、A_2V_1、A_3V_1、A_4V_1、A_1V_2、A_1V_3、A_1V_4	声音刺激: 80 dB 声压级，呈现 13 ms; 闪光刺激: 呈现在背景为黑色，中央注视点下方 5°，直径 2°，亮度 118 cd/m^2 的白色闪光，呈现 13 ms; 时间间隔: SOA $_{(A1V1)}$ = 26 ms; SOA $_{(V1V2)}$ = 65 ms; SOA $_{(A1A2)}$ = 52 ms。	通过 tDCS 上调或下调皮质兴奋性可以调节视错觉，颞、枕皮质区(后顶叶皮层区除外)与闪光错觉存在功能相关
Kamke 等 (2012)	裂变错觉	TMS	V_1、V_2、A_2V_1、A_2V_1	声音刺激: 3.5 kHz 频率, 80 dB 声压级，呈现 8 ms; 闪光刺激：呈现在背景为黑色，中央注视点左侧或右侧 8°，直径为 2°的白色闪光，呈现 20 ms; 时间间隔: SOA $_{(V1V2)}$ = 70 ms; SOA $_{(V1A2)}$ = 70 ms; SOA $_{(V1A2)}$ = 160 ms	右侧角回改变了刺激驱动的注意力网络，降低了视听刺激错误整合的可能性
Shams et al., (2005)	闪光错觉	MEG	V_1、V_2、V_3、V_4、A_1、A_2、A_3、A_4、A_1V_1、A_1V_2、A_1V_3、A_1V_4、A_2V_1、A_2V_2、A_2V_3、A_2V_4、A_3V_1、A_3V_2、A_3V_3、A_3V_4、A_4V_1、A_4V_2、A_4V_3、A_4V_4	声音刺激: 3.5 kHz 频率, 80 dB 声压级，呈现 10 ms; 闪光刺激：呈现在背景为黑色，中央注视点下方 12°，直径 1.5°的白色闪光，呈现 10 ms; 时间间隔: SOA $_{(V1V2)}$ = 70 ms; SOA $_{(A1A2)}$ = 58 ms	枕顶叶皮层区域在视觉刺激开始的 35~65 ms 受到了调节，在刺激后约 150 ms，枕叶、顶叶以及前部区域的活动也被调节
Keil et al., (2013)	闪光错觉	MEG	V_1、V_2、A_1、A_2、A_1V_1、A_1V_2、A_2V_1、A_2V_2	声音刺激：呈现 7 ms; 闪光刺激：呈现 17 ms; 时间间隔: ISI $_{(V1V2)}$ = 50 ms; SOA $_{(A1A2)}$ = 64 ms; SOA $_{(V1V2)}$ = 67 ms。	在产生错觉之前，左侧颞中回和听觉区域之间的 β 波段相位同步性增强，而视觉区域的相位同步性减弱
Chan et al., (2017)	裂变错觉	MEG	V_1、V_2、A_1、A_2、A_1V_1、A_2V_1、A_2V_2	声音刺激: 3.5 kHz 频率, 65 dB 声压级，呈现 16 ms; 闪光刺激：呈现在背景为黑色，中央注视点下方 4°，直径 2°的白色闪光，呈现 16 ms; 时间间隔: SOA = 50~100 ms。	老年人对将要呈现的刺激产生了预测，视听皮层的跨通道连接对与错觉相关的调节能力更强

不统一，因此研究结果之间也存在一些差异。未来的研究应该整合相关研究进一步考察声音诱发闪光错觉的神经机制，并且可以在传统分析的基础上，结合功能连通性的分析，以及结合静息态和结构像的数据进行相关预测性分析来进一步探讨声音诱发闪光产生的机制。

4 总结与展望

声音诱发闪光错觉效应反映了视听觉信息整合过程中听觉信息对视觉信息感知的影响。本文对声音诱发闪光错觉效应的研究进行了系统性梳理，主要包括：(1)声音诱发闪光错觉被试内差异的影响因素，包括自下而上的物理刺激因素(刺激空间特性、刺激强度差异和刺激间的时间间隔)和自上而下的认知因素(注意分配、内源性注意、反馈方式和认知期望)等，还包括视听刺激依赖程度、视听刺激整合的发展以及视听刺激知觉敏感性等造成被试间差异的因素。(2)声音诱发闪光错觉的认知神经机制，包括采用 EEG、ERP 和 MEG 等手段考察的视听整合过程的早期加工阶段，以及采用 MRI、tDCS、TMS 和 MEG 等手段发现的枕叶视觉皮层、颞上沟、上丘、前额叶和小脑等广泛加工声音诱发闪光错觉的脑区。自声音诱发闪光错觉效应被发现以来，研究者们开展了大量的研究。通过对本文的综述，不但能够系统地了解声音诱发闪光错觉效应及其机制，同时也能够为视听整合类的相关研究提供借鉴。

目前声音诱发闪光错觉的研究还存在一些局限和不足。比如目前仅限于自身机制探讨，即一方面是影响因素类研究，一方面是神经机制类研究。尚待解决的是将声音诱发闪光错觉的研究与其他视听整合类的研究进行比较，探究声音诱发闪光错觉和其他视听整合类研究的共性问题，以及声音诱发闪光错觉相对于其他视听整合类研究的特异性的问题，为视听整合甚至多感觉整合的相关研究提供更多实证性的支持。下面将结合已有研究的不足，提出未来研究的可能性的方向。

第一，注意对声音诱发闪光错觉的影响。未来的研究可以在实验研究中系统地调节不同刺激的竞争程度，例如通过操纵知觉负载操纵自上而下的注意，考察其对多感觉整合产生的影响(Talsma, Senkowski, Woldorff, & Woldorff, 2010)。也可以在现有多感觉整合与注意交互作用研究的基础上(Tang, Wu, & Shen, 2016)，探讨不同注意模式下的声音诱发闪光错觉效应。一方面为影响声音诱发闪光错觉在被试内差异的因素提供更多来自注意这一因素的实证研究，另一方面也能够为视听整合类的研究提供来自不同感觉通道权重不同情况下产生的感觉主导效应的证据支持。

第二，奖赏对声音诱发闪光错觉的影响。Rosenthal 等人(2009)采用 Shams 等人(2002)的研究范式发现，在给予被试货币奖励时可以降低被试知觉到的错觉闪光，但具体奖励机制为什么会降低错觉知觉，没有过多研究。为什么奖励可以改变被试知觉?是否和奖励机制的神经基础有关？那么高奖励和低奖励是否对被试知觉错觉有不同影响？未来研究可以通过操纵奖赏的量化等方式进一步考察奖赏对声音诱发闪光错觉的影响，一方面为影响声音诱发闪光错觉在个体内变异的因素提供更多来自奖赏这一因素的实证研究，另一方面也能够为奖赏如何影响视听整合加工提供来自听觉主导效应方向的证据支持。

第三，视听整合的方式对声音诱发闪光错觉的影响。以往关于声音诱发闪光错觉效应的研究均认为错觉效应的出现是由于视听刺激在整合的过程中大脑对视觉通道和听觉通道分配的权重不一致，导致整合的过程中出现了听觉通道信息的加工主导了视觉通道信息的加工位(Zhou, Jiang, He, & Chen, 2010)。但到底声音诱发闪光错觉效应是由于整合过程中出现了感觉通道的主导，还是由于在视觉通道信息和听觉通道信息在输入的过程中各个通道中信息加工的差异导致的，目前还不能确定。未来研究可以关注单通道的视觉刺激和单通道的听觉刺激的加工在声音诱发闪光错觉中所起到的作用，一方面能够明确在声音诱发闪光错觉中视觉与听觉输入的信息是先得到整合以后才在脑区进一步加工，还是在进入脑区之前单通道就已经加工对应的信息然后进一步整合的。另一方面也能够考察声音诱发闪光错觉这种视听整合现象与其他视听整合现象的共性与特性。

第四，声音诱发闪光错觉对记忆和学习等认知过程的影响。目前声音诱发闪光错觉的研究主要集中在影响因素和神经机制层面的探讨，很少关注声音诱发闪光错觉效应如何影响后续的认知加工过程。例如，Quak, London 和 Talsma (2015)认为多感觉通道信息加工的方式能够很好地理解

工作记忆如何存储和操纵信息(Quak et al., 2015)。也有一系列的相关研究对此问题进行了探究，证实了多感觉整合加工过程有助于工作记忆的编码和提取(Barutchu, Sahu, Humphreys, & Spence, 2019; Brunetti, Indraccolo, Mastroberardino, Spence, & Santangelo, 2017; Xie, Xu, Bian, & Li, 2017)。因此，未来研究可以考察声音诱发闪光错觉这种听觉主导效应类的视听整合对工作记忆以及学习的影响，一方面能够考察声音诱发闪光错觉这种视听整合现象与其他视听整合现象的共性与特性。另一方面也能够扩展声音诱发闪光错觉的研究角度和思路。

第五，应用计算模型探讨声音诱发闪光错觉的神经机制。以往的研究从计算模型的角度对视听整合中经典视觉主导效应(McGurk效应)机制进行了解释(Marques, Lapenta, Costa, & Boggio, 2016; Samuel, 2011)，比如分层预测编码模型(Olasagasti, Bouton, & Giraud, 2015)以及差异噪声编码模型(noisy encoding of disparity model, NED) (Magnotti & Beauchamp, 2015)。在应用计计算模型的过程中尝试先假定其中的加工过程，并利用不同的参数来描述不同的加工过程，参数在其中代表的意义与特定加工过程相对应(罗霄骁，康冠兰，周晓林，2018)。但目前很少有声音诱发闪光错觉效应的研究将神经生理技术与计算模型相结合。现有的一项关于声音诱发闪光错觉的计算模型研究通过一个由两层视觉和听觉神经元组成的简单神经网络考察声音诱发闪光错觉中的视听交互过程(Cuppini, Magosso, Bolognini, Vallar, & Ursino, 2014)。未来的研究可以关于声音诱发闪光错觉效应的计算模型研究可以考虑结合分层预测编码模型和差异噪声编码模型，利用 EEG/ERP、fMRI 结果等神经科学指标进行参数拟合，或是神经科学研究可以考虑利用计算模型寻找相应参数的对应脑区，为模型的参数找到神经基础。

参考文献

罗霄骁，康冠兰，周晓林. (2018). McGurk 效应的影响因素与神经基础. *心理科学进展, 26*(11), 1935–1951.

文小辉，李国强，刘强. (2011). 视听整合加工及其神经机制. *心理科学进展, 19*(7), 976–982.

于薇，王爱君，张明. (2017). 集中和分散注意对多感觉整合中听觉主导效应的影响. *心理学报, 49*(2), 164–173.

张明，唐晓雨，于薇，宁波，王智楠，王爱君. (2018). 基于通道的内源性注意对声音诱发闪光错觉的影响. *心理学报, 50*(11), 20–29.

Abadi, R. V., & Murphy, J. S. (2014). Phenomenology of the sound-induced flash illusion. *Experimental Brain Research, 232*(7), 2207–2220.

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research, 21*(3), 301–308.

Balz, J., Keil, J., Romero, Y. R., Mekle, R., Schubert, F., Aydin, S., ... Senkowski, D. (2016). GABA concentration in superior temporal sulcus predicts gamma power and perception in the sound-induced flash illusion. *NeuroImage, 125*, 724–730.

Balz, J., Roa Romero, Y., Keil, J., Krebber, M., Niedeggen, M., Gallinat, J., & Senkowski, D. (2016). Beta/Gamma oscillations and event-related potentials indicate aberrant multisensory processing in schizophrenia. *Frontiers in Psychology, 7*, 1896.

Bao, V. A., Doobay, V., Mottron, L., Collignon, O., & Bertone, A. (2017). Multisensory integration of low-level information in autism spectrum disorder: Measuring susceptibility to the flash-beep illusion. *Journal of Autism and Developmental Disorders, 47*(8), 2535–2543.

Barutchu, A., Sahu, A., Humphreys, G. W., & Spence, C. (2019). Multisensory processing in event-based prospective memory. *Acta Psychologica, 192*, 23–30.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension: Evidence from auditory–visual temporal order judgment. *International Journal of Psychophysiology, 50*(1–2), 147–155.

Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding: Experience-dependent effects in the double-flash illusion. *Experimental Brain Research, 234*(10), 3037–3047.

Bolognini, N., Rossetti, A., Casati, C., Mancini, F., & Vallar, G. (2011). Neuromodulation of multisensory perception: A tDCS study of the sound-induced flash illusion. *Neuropsychologia, 49*(2), 231–237.

Brighina, F., Bolognini, N., Cosentino, G., Maccora, S., Paladino, P., Baschi, R., ... Fierro, B. (2015). Visual cortex hyperexcitability in migraine in response to sound-induced flash illusions. *Neurology, 84*(20), 2057–2061.

Brunetti, R., Indraccolo, A., Mastroberardino, S., Spence, C., & Santangelo, V. (2017). The impact of cross-modal correspondences on working memory performance. *Journal of Experimental Psychology: Human Perception and Performance, 43*(4), 819–831.

Cecere, R., Rees, G., & Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception.

Current Biology, 25(2), 231–235.

Chan, J. S., Connolly, S. K., & Setti, A. (2018). The number of stimulus-onset asynchronies affects the perception of the sound-induced flash illusion in young and older adults. *Multisensory Research, 31*(3–4), 175–190.

Chan, J. S., Kaiser, J., Brandl, M., Matura, S., Prvulovic, D., Hogan, M., & Naumer, M. J. (2015). Expanded temporal binding windows in people with mild cognitive impairment. *Current Alzheimer Research, 12*(1), 61–68.

Chan, J. S., Wibral, M., Wollstadt, P., Stawowsky, C., Brandl, M., Helbling, S., ... Kaiser, J. (2017). Predictive coding over the lifespan:Increased reliance on perceptual priors in older adults — A magnetoencephalography and dynamic causal modelling study. *BioRxiv*, 178095.

Cuppini, C., Magosso, E., Bolognini, N., Vallar, G., & Ursino, M. (2014). A neurocomputational analysis of the sound-induced flash illusion. *NeuroImage*, *92*, 248–266.

de Haas, B., Kanai, R., Jalkanen, L., & Rees, G. (2012). Grey matter volume in early human visual cortex predicts proneness to the sound-induced flash illusion. *Proceedings of the Royal Society B: Biological Sciences, 279*(1749), 4955–4961.

DeLoss, D. J., & Andersen, G. J. (2015). Aging, spatial disparity, and the sound-induced flash illusion. *PLoS ONE, 10*(11), e0143773.

DeLoss, D. J., Pierce, R. S., & Andersen, G. J. (2013). Multisensory integration, aging, and the sound-induced flash illusion. *Psychology Aging, 28*(3), 802–812.

Georgios, M., & Julian, K. (2018). High cognitive load enhances the susceptibility to non-speech audiovisual illusions. *Scientific Reports, 8*(1), 11530.

Gieseler, A., Tahden, M. A. S., Thiel, C. M., & Colonius, H. (2018). Does hearing aid use affect audiovisual integration in mild hearing impairment? *Experimental Brain Research, 236,* 1161–1179.

Haß, K., Sinke, C., Reese, T., Roy, M., Wiswede, D., Dillo, W., ... Szycik, G. R. (2017). Enlarged temporal integration window in schizophrenia indicated by the double-flash illusion. *Cognitive Neuropsychiatry, 22*(2), 145–158.

Hernández, B., Setti, A., Kenny, R. A., & Newell, F. N. (2019). Individual differences in ageing, cognitive status, and sex on susceptibility to the sound-induced flash illusion: A large-scale study. *Psychology and Aging, 34*(7), 978–990.

Innes-Brown, H., & Crewther, D. (2009). The impact of spatial incongruence on an auditory-visual illusion. *PLoS ONE, 4*(7), e6450.

Jiang, Y., & Han, S. H. (2007). Perceptual gain and perceptual loss: Distinct neural mechanisms of audiovisual interactions. *International Journal of Magnetic Resonance Imaging*, *1*(1), 1749–8023.

Kaiser, M., Senkowski, D., Busch, N. A., Balz, J., & Keil, J. (2019). Single trial prestimulus oscillations predict perception of the sound-induced flash illusion. *Scientific Reports, 9*(1), 5983.

Kamke, M. R., Vieth, H. E., Cottrell, D., & Mattingley, J. B. (2012). Parietal disruption alters audiovisual binding in the sound-induced flash illusion. *NeuroImage, 62*(3), 1334–1341.

Kawabe, T. (2009). Audiovisual temporal capture underlies flash fusion. *Experimental Brain Research, 198*(2–3), 195–208.

Keil, J., Müller, N., Hartmann, T., & Weisz, N. (2013). Prestimulus beta power and phase synchrony influence the sound-induced flash illusion. *Cerebral Cortex, 24*(5), 1278–1288.

Keil, J., & Senkowski, D. (2017). Individual alpha frequency relates to the sound-induced flash illusion. *Multisensory Research, 30*(6), 565–578.

Kim, R., Peters, M. A. K., & Shams, L. (2011). 0 + 1 > 1: How adding noninformative sound improves performance on a visual task. *Psychological Science*, *23*(1), 6–12.

Klatzky, R. L., Marston, J. R., Giudice, N. A., Golledge, R. G., & Loomis, J. M. (2006). Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of Experimental Psychology: Applied, 12*(4), 223–232.

Kostaki, M., & Vatakis, A. (2016). Crossmodal binding rivalry: A “race” for integration between unequal sensory inputs. *Vision Research, 127*, 165–176.

Kumpik, D. P., Roberts, H. E., King, A. J., & Bizley, J. K. (2014). Visual sensitivity is a stronger determinant of illusory processes than auditory cue parameters in the sound-induced flash illusion. *Journal of Vision*, *14*(7), 12.

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging, 27*(8), 1155–1163.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16*(3), 468–478.

Maccora, S., Giglia, G., Bolognini, N., Cosentino, G., Gangitano, M., Salemi, G., & Brighina, F. (2019). Cathodal occipital tDCS is unable to modulate the sound induced flash illusion in migraine. *Frontiers in Human Neuroscience, 13*, 247.

Magnotti, J. F., & Beauchamp, M. S. (2015). The noisy encoding of disparity model of the McGurk effect. *Psychonomic Bulletin & Review*, *22*(3), 701–709.

Mahoney, J. R., Li, P. C. C., Oh-park, M., Verghese, J., &

Holtzer, R. (2011). Multisensory integration across the senses in young and old adults. *Brain Research, 1426*(17), 43–53.

Marques, L. M., Lapenta, O. M., Costa, T. L., & Boggio, P. S. (2016). Multisensory integration processes underlying speech perception as revealed by the McGurk illusion. *Language, Cognition and Neuroscience*, *31*(9), 1115–1129.

Máté, A., Anette, G., Verena, C., Masataka, W., & Uta, N. (2015). A spatially collocated sound thrusts a flash into awareness. *Frontiers in Integrative Neuroscience, 9*, 1–8.

McCormick, D., & Mamassian, P. (2008). "What does the illusory-flash look like?". *Vision Research, 48*(1), 63–69.

Mcgovern, D. P., Roudaia, E., Stapleton, J., McGinnity, T. M., & Newell, F. N. (2014). The sound-induced flash illusion reveals dissociable age-related effects in multisensory integration. *Frontiers in Aging Neuroscience, 6, 250*.

Merriman, N. A., Whyatt, C., Setti, A., Craig, C., & Newell, F. N. (2015). Successful balance training is associated with improved multisensory function in fall-prone older adults. *Computers in Human Behavior, 45*, 192–203.

Mishra, J., Martinez, A., & Hillyard, S. A. (2008). Cortical processes underlying sound-induced flash fusion. *Brain Research, 1242*(4), 102–115.

Mishra, J., Martínez, A., & Hillyard, S. A. (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *Journal of Cognitive Neuroscience, 22*(8), 1714–1729.

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience, 27*(15), 4120–4131.

Moro, S. S., & Steeves, J. K. E. (2018). Normal temporal binding window but no sound-induced flash illusion in people with one eye. *Experimental Brain Research, 236*(6), 1825–1834.

Narinesingh, C., Goltz, H. C., & Wong, A. M. F. (2017). Temporal binding window of the sound-induced flash illusion in amblyopia. *Investigative Opthalmology & Visual Science, 58*(3), 1442–1448.

O'Brien, J., Ottoboni, G., Tessari, A., & Setti, A. (2017). One bout of open skill exercise improves cross-modal perception and immediate memory in healthy older adults who habitually exercise. *PLoS ONE, 12*(6), e0178739.

Olasagasti, I., Bouton, S., & Giraud, A. L. (2015). Prediction across sensory modalities: A neurocomputational model of the McGurk effect. *Cortex*, *68*, 61–75.

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, *9*, 197.

Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brain Topography, 21*(3–4), 185–192.

Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology*, *62*(1), 49–72.

Setti, A., & Chan, J. S. (2011). Familiarity of objects affects susceptibility to the sound-induced flash illusion. *Neuroscience Letters, 492*(1), 19–22.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425–432.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature, 408*(6814), 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research, 14*(1), 147–152.

Shams, L., Kamitani, Y., Thompson, S. M., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport, 12*(17), 3849–3852.

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport, 16*(17), 1923–1927.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266.

Sterzer, P., Kleinschmidt, A., & Rees, G. (2009). The neural bases of multistable perception. *Trends in Cognitive Sciences*, *13*(7), 310–318.

Sun, Y. W., Liu, X. L., Li, B. L., Sava-Segal, C., Wang, A., & Zhang, M. (2020). Effects of repetition suppression on sound induced flash illusion with aging. *Frontiers in Psychology*. *11*, 216.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14*(9), 400–410.

Tang, X. Y., Wu, J. L., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience & Biobehavioral Reviews*, *61*, 208–224.

Wang, A. J., Sang, H. B., He, J. Y., Sava-Segal, C., Tang, X. Y., & Zhang, M. (2019). Effects of cognitive expectation on sound-induced flash illusion. *Perception*, *48*(12), 1214–1234.

Wang, M., Arteaga, D., & He, B. J. (2013). Brain mechanisms for simple perception and bistable perception. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(35), E3350–3359.

Watkins, S., Shams, L., Josephs, O., & Rees, G. (2007). Activity in human V1 follows multisensory perception. *NeuroImage, 37*(2), 572–578.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association

with illusory visual perception. *NeuroImage, 31*(3), 1247–1256.

Whittingham, K. M., Mcdonald, J. S., & Clifford, C. W. G. (2014). Synesthetes show normal sound-induced flash fission and fusion illusions. *Vision Research, 105*, 1–9.

van Erp, J. B. F., Philippi, T. G., & Werkhoven, P. (2013). Observers can reliably identify illusory flashes in the illusory flash paradigm. *Experimental Brain Research, 226*(1), 73–79.

Vanes, L. D., White, T. P., Wigton, R. L., Joyce, D., Collier, T., & Shergill, S. S. (2016). Reduced susceptibility to the sound-induced flash fusion illusion in schizophrenia. *Psychiatry Research, 245*, 58–65.

Xie, Y. J., Xu, Y. Y., Bian, C., & Li, M. (2017). Semantic congruent audiovisual integration during the encoding stage of working memory: An ERP and sLORETA study. *Scientific Reports*, *7*(1), 5112.

Yalachkov, Y., Bergmann, H. J., Soydaş, D., Buschenlange, C., Fadai Motlagh, L. Y., Naumer, M. J., ... Gehrig, J. (2019). Cognitive impairment in multiple sclerosis is reflected by increased susceptibility to the sound-induced flash illusion. *Frontiers in Neurology, 10*, 373.

Zhang, N. Y., & Chen, W. (2006). A dynamic fMRI study of illusory double-flash effect on human visual cortex. *Experimental Brain Research, 172*(1), 57–66.

Zhou, W., Jiang, Y., He, S., & Chen, D. (2010). Olfaction modulates visual perception in binocular rivalry. *Current Biology*, *20*(15), 1356–1358.

Sound-induced flash illusion in multisensory integration

WANG Aijun[1]; HUANG Jie[1]; LU Feifei[1]; HE Jiaying[1]; TANG Xiaoyu[2]; ZHANG Ming[1]

([1] *Department of Psychology, Soochow University, Research Center for Psychology and Behavioral Sciences, Suzhou* 215123, *China*) ([2] *School of Psychology, Liaoning Normal University, Liaoning Collaborative Innovation Center of Children and Adolescents Healthy Personality Assessment and Cultivation, Dalian* 116029, *China*)

Abstract: Sound-induced Flash Illusion (SiFI) is an auditory-dominated multisensory integration phenomenon that occurs when flashes presented in conjunction with an incongruent number of auditory sounds results in the flashes being perceived as equal in number to the auditory sounds. Here we summarize several factors that could impact on SiFI: (1) Bottom-up variance factors comprising physical properties (i.e., spatial characteristics of stimuli, differences in stimulus intensity, and time interval between audiovisual stimuli), top-down variance factors comprisingcognitive processes (i.e., attention, feedback patterns, and cognitive expectations), and between-subjects variance factors, such as the dependence for audiovisual stimuli, the ability to integrate audiovisual stimuli, and perceptual sensitivity to audiovisual stimuli; (2) Cognitive processing stages and neural mechanisms of SiFI (i.e., related brain regions). Future studies should focus on the effects of cognitive processes such as attention, reward, audiovisual integration on SiFI. The effects of SiFI on memory and learning should also be explored in future research. In addition, future studies combining computational models and neuroscience should further explore the cognitive and neural mechanisms of SiFI.

Key words: sound induced flash illusion (SiFI); fission illusion; fusion illusion; multisensory integration